Simultaneous Confidence Bands for Penalized Spline Estimators

Tatyana Krivobokova[1]
Georg August Universität Göttingen

Thomas Kneib[2]
Carl von Ossietzy Universität Oldenburg

Gerda Claeskens[3]
Katholieke Universiteit Leuven

18th March 2010

Abstract

In this paper we construct simultaneous confidence bands for a smooth curve using penalized spline estimators. We consider three types of estimation methods: (i) as a standard (fixed effect) nonparametric model, (ii) using the mixed model framework with the spline coefficients as random effects and (iii) a full Bayesian approach. The volume-of-tube formula is applied for the first two methods and compared from a frequentist perspective to Bayesian simultaneous confidence bands. It is shown that the mixed model formulation of penalized splines can help to obtain, at least approximately, confidence bands with either Bayesian or frequentist properties. Simulations and data analysis support the methods proposed. The R package *ConfBands* accompanies the paper.

Key words and phrases. Bayesian penalized splines; B-splines; Confidence band; Mixed model; Penalization.

[1]CRC PEG, Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany
[2]Mathematics Department, Carl von Ossietzy Universität Oldenburg, 26111 Oldenburg, Germany
[3]ORSTAT, K.U. Leuven, Naamsestraat 69, 3000 Leuven, Belgium.

1 Introduction

Penalized spline smoothing has received much attention over the last decade. Eilers and Marx (1996) coined the term "P-spline" estimator for a version of the O'Sullivan (1986) estimator with a simplified penalty matrix. The idea is to estimate the function of interest by some spline. Thereby a generous basis dimension is taken and penalization with an integrated squared derivative of the spline function helps to avoid overfitting. A small parameter dimension, a flexible choice of basis and penalties, and direct links to mixed and Bayesian models made this smoothing technique popular, see Ruppert et al. (2003) for examples and applications.

The theoretical properties of penalized splines remained less explored. Some first result can be found in Hall and Opsomer (2005), Li and Ruppert (2008) and Kauermann et al. (2009). Recently Claeskens et al. (2009) showed that depending on the number of knots, the asymptotic scenario of the penalized spline estimator is similar to that of either regression spline or smoothing spline estimator. Thereby the optimal asymptotic orders for the number of spline functions and for the smoothing parameter were obtained. These new results can now be applied for inference, in particular for the construction of simultaneous confidence bands.

In general, simultaneous confidence bands for a function f are constructed by studying the asymptotic distribution of $\sup_{a\le x\le b}|\widehat{f}(x)-f(x)|$. The approach by Bickel and Rosenblatt (1973) relates this to a study of the distribution of $\sup_{a\le x\le b}|Z(x)|$, with $Z(x)$ a (standardized) Gaussian process satisfying certain conditions, which they show to have an asymptotic extreme value distribution. This approach for the construction of confidence bands has been used in the context of nonparametric estimation by, amongst others, Härdle (1989) for M-estimators and Claeskens and Van Keilegom (2003) for local polynomial likelihood estimators. Hall (1991) studied the convergence of normal extremes and found them to be slow, with the consequence that all those confidence bands do not perform satisfactorily for small samples, and bootstrap methods are often applied (see for example, Neumann and Polzehl, 1998; Claeskens and Van Keilegom, 2003).

Knafl et al. (1985) and Hall and Titterington (1988) developed confidence bands based on large-sample upper bounds for the size of $\sup_{a\leq x\leq b}|\widehat{f}(x)-f(x)|$. The main challenge of this approach is to take the bias of a nonparametric estimator into account. Also the choice of the smoothing parameter is a delicate matter. Eubank and Speckman (1993) applied a similar technique to obtain confidence bands for a periodic twice differentiable function, using a kernel estimator. Thereby the smoothing parameter was chosen data-driven and the bias was approximated using the estimator of the second derivative of the underlying mean function. Xia (1998) extended the approach of Eubank and Speckman (1993) using local polynomial estimators.

Another attractive approach is to construct confidence bands based on the volume of tube formula. Sun (1993) studied the tail probabilities of suprema of Gaussian random processes, which can be used for the construction of simultaneous confidence bands. It turns out that the leading coefficient in the approximation of the tail probability $P(\sup_{a\leq x\leq b}|Z(x)|>c)$ for $c\to\infty$ is connected through Weyl's (1939) formula for the volume of a tube of a manifold (also referred to as a Hotelling (1939) formula) to the volume of the manifold embedded in a unit sphere. The main attraction of this method is its straightforward extendability to more general and high dimensional settings. However, the problem of the smoothing parameter choice and handling the bias still remains an important issue. Sun and Loader (1994) suggested a bias correction for a particular class of functions, but left the smoothing parameter choice open. Zhou et al. (1998, Theorem 4.2) used the volume-of-tube formula for estimation by regression splines (without using a penalty), but did not account for the bias, which lead to undercoverage. We will use this method for the construction of confidence bands for estimation by penalized spline estimators in the fixed and mixed model framework.

In contrast to the frequentist setting, the Bayesian confidence bands are constructed based on the posterior distribution of the underlying process, given the data. Even though highest posterior density credible bands should be optimal from a theoretical perspective, they are in general hard to obtain, in particular when the estimation is based on Markov chain Monte Carlo (MCMC) simulation techniques (as will be the case in this paper and is common practice in complex statistical models). In this case, the posterior density is not available and, as a consequence, confidence intervals are typically constructed based on sample quantiles obtained from the Monte Carlo output. The difficulty in constructing simultaneous confidence bands then lies in combining the sample quantiles such that a simultaneous coverage for a vector parameter is achieved. Besag et al. (1995) propose to combine appropriate order statistics of the univariate samples. Crainiceanu et al. (2007) consider simultaneous confidence bands when posterior normality for the parameter vector can be assumed. Held (2004) constructs simultaneous posterior probability statements about vector parameters based on a Rao-Blackwellized estimate of the posterior density. In principle, the posterior probabilities could be inverted to obtain a highest posterior density credible band, but the computational burden is high since additional simulations are required to obtain the posterior density estimate.

In this work we demonstrate advantages of the mixed model formulation, which combines both frequentist and Bayesian approaches. We develop a new approach for the mixed model based confidence bands, as well as a new Bayesian simultaneous confidence band. In fact, the confidence bands obtained in the marginal mixed model framework are identical to the Bayesian ones, up to an unaccounted variability due to variance estimation. Since the Bayesian confidence bands (and thus the marginal mixed model based ones) tend to be conservative in the nonparametric setting (see Cox, 1993), we show how the confidence bands with the approximately frequentist properties can be obtained using the mixed model representation of penalized splines. Thereby no explicit bias estimation is necessary and the smoothing parameter is estimated in the usual way from the corresponding (restricted) likelihood.

We first introduce the curve estimators in Section 2, then, in Sections 3 and 4 we construct confidence bands for each setting, where we obtain a new result for the mixed models as well as for the Bayesian method. A comparison and discussion follows in Section 5, while simulation results and a data example are contained in Sections 6 and 7.

2 Penalized splines in three frameworks

We wish to construct a simultaneous confidence band for an unknown smooth function $f \in C^q([a,b])$, which is a q times continuously differentiable function. We have observations (Y_i, x_i), with $x_i \in [a,b]$, $i = 1, \ldots, n$, from the model

$$Y_i = f(x_i) + \varepsilon_i. \tag{1}$$

The residuals ε_i are assumed to be independent and identically distributed as $N(0, \sigma_\epsilon^2)$. We first introduce some notation and explain the three frameworks for penalized splines.

2.1 Penalized spline estimator

We denote by $S(p+1; \underline{\tau})$ the set of spline functions of degree p with knots $\underline{\tau} = \{a = \tau_0 < \tau_1 < \ldots < \tau_K < \tau_{K+1} = b\}$. This set consists of all functions that are a polynomial of degree p on each interval $[\tau_j, \tau_{j+1}]$, and are $p-1$ times continuously differentiable. The set $S(1, \underline{\tau})$ consists of piecewise constant functions with jumps at the knots.

A penalized spline estimator of degree p based on the set of knots $\underline{\tau}$ is the solution to

$$\min_{s(x) \in S(p+1;\underline{\tau})} \left[\sum_{i=1}^{n} \{Y_i - s(x_i)\}^2 + \lambda \int_a^b \left\{ s^{(q)}(x) \right\}^2 dx \right], \tag{2}$$

with $q \leq p$. Denote by a row vector $\boldsymbol{P}(x, \underline{\tau}) = \{P_1(x, \underline{\tau}), \ldots, P_{K+p+1}(x, \underline{\tau})\}$ a basis for $S(p+1, \underline{\tau})$. One example is the set of polynomial and piecewise polynomial functions $\{1, x, \ldots, x^p, (x-\tau_1)_+^p, \ldots, (x-\tau_K)_+^p\}$, another example is a basis of B-spline functions of degree p. With this notation, the spline function can be written as $s(x) = \boldsymbol{P}(x, \underline{\tau})\boldsymbol{\theta}$, with an unknown parameter $\boldsymbol{\theta}$ of length $K+p+1$ and (2) can be represented as minimization problem over $\boldsymbol{\theta}$.

The penalty in (2) is the integrated squared qth derivative of the spline function, which is assumed to be finite. Let $\boldsymbol{D}$ be the matrix such that $\int_a^b \left[\{\boldsymbol{P}(x, \underline{\tau})\boldsymbol{\theta}\}^{(q)}\right]^2 dx = \boldsymbol{\theta}^t \boldsymbol{D} \boldsymbol{\theta}$. Define the spline basis matrix $\boldsymbol{P} = \{\boldsymbol{P}(x_1, \underline{\tau})^t, \ldots, \boldsymbol{P}(x_n, \underline{\tau})^t\}^t$, and the response vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^t$, then, for a given λ, the penalized spline estimator can be written as

$$\widetilde{\boldsymbol{f}} = \boldsymbol{P}\widetilde{\boldsymbol{\theta}} = \boldsymbol{P}(\boldsymbol{P}^t\boldsymbol{P} + \lambda\boldsymbol{D})^{-1}\boldsymbol{P}^t\boldsymbol{Y}, \tag{3}$$

where the estimator $\widetilde{\boldsymbol{f}} = \{\tilde{f}(x_1), \ldots, \tilde{f}(x_n)\}^t$.

The penalty constant λ plays the role of a smoothing parameter. It can be estimated with any data-driven method that asymptotically minimizes the average mean squared error, like (generalized) cross validation or the Akaike information criterion (AIC). Replacing λ by its estimate $\hat{\lambda}$, one gets $\widehat{\boldsymbol{f}} = \boldsymbol{P}(\boldsymbol{P}^t\boldsymbol{P} + \hat{\lambda}\boldsymbol{D})^{-1}\boldsymbol{P}^t\boldsymbol{Y}$.

2.2 Penalized spline estimators as predictors in mixed models

A penalized spline estimator is equivalent to a best linear unbiased predictor (BLUP) in the corresponding mixed model (Brumback et al., 1999). To show this, we first decompose

$$\boldsymbol{P}\boldsymbol{\theta} = \boldsymbol{P}(\boldsymbol{F}_\beta\boldsymbol{\beta} + \boldsymbol{F}_u\boldsymbol{u}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \tag{4}$$

such that $(\boldsymbol{F}_\beta, \boldsymbol{F}_u)$ is of full rank, providing uniqueness of transformation, and $\boldsymbol{F}_\beta^t\boldsymbol{F}_u = \boldsymbol{F}_u^t\boldsymbol{F}_\beta = \boldsymbol{F}_\beta^t\boldsymbol{D}\boldsymbol{F}_\beta = 0$, $\boldsymbol{F}_u^t\boldsymbol{D}\boldsymbol{F}_u = \boldsymbol{I}_{K+p+1-q}$, ensuring that only coefficients $\boldsymbol{u}$ are penalized. Thereby $\boldsymbol{P}$ is a $n \times (K+p+1)$, $\boldsymbol{X}$ is a $n \times q$ and $\boldsymbol{Z}$ is a $n \times \widetilde{K}$ matrix, with $\widetilde{K} = K+p+1-q$. There are several approaches to obtain such a decomposition, for more details consult e.g. Durban and Currie (2003) or Fahrmeir et al. (2004). If we assume that $\boldsymbol{Y}|\boldsymbol{u} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \sigma_\epsilon^2 \boldsymbol{I}_n)$ and $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{\widetilde{K}})$, this leads to the standard linear mixed model with the best linear unbiased predictor (BLUP)

$$\tilde{\boldsymbol{f}}_m = \boldsymbol{P}_m \tilde{\boldsymbol{\theta}}_m = \boldsymbol{P}_m \left(\boldsymbol{P}_m^t \boldsymbol{P}_m + \frac{\sigma_\epsilon^2}{\sigma_u^2} \boldsymbol{D}_m \right)^{-1} \boldsymbol{P}_m^t \boldsymbol{Y},$$

where $\boldsymbol{P}_m = [\boldsymbol{X}, \boldsymbol{Z}]$, $\boldsymbol{\theta}_m = [\boldsymbol{\beta}, \boldsymbol{u}]$, $\boldsymbol{D}_m = \text{diag}\{\boldsymbol{0}_q, \boldsymbol{1}_{\widetilde{K}}\}$.

If we replace further σ_ϵ^2 and σ_u^2 with the corresponding (restricted) maximum likelihood estimators in the mixed model, this results in the estimated best linear unbiased predictor (EBLUP) $\widehat{\boldsymbol{f}}_m = \boldsymbol{P}_m \widehat{\boldsymbol{\theta}}_m = \boldsymbol{P}_m \left(\boldsymbol{P}_m^t \boldsymbol{P}_m + \hat{\sigma}_\epsilon^2 / \hat{\sigma}_u^2 \boldsymbol{D}_m \right)^{-1} \boldsymbol{P}_m^t \boldsymbol{Y}$. Due to the construction of $\boldsymbol{P}_m$ there always exists a square invertible matrix $\boldsymbol{L}$ such that $\boldsymbol{P} = \boldsymbol{P}_m \boldsymbol{L}$ and $\boldsymbol{D} = (\boldsymbol{L}^{-1})^t \boldsymbol{D}_m \boldsymbol{L}^{-1}$. We therefore do not further distinguish between the different forms for the model and penalty matrices. However, the notation with a subscript 'm' as in $\widehat{\boldsymbol{f}}_m$ and $\widehat{\boldsymbol{\theta}}_m$ will stress that that the estimators are obtained in the mixed model framework. Note that the smoothing parameter in this mixed model formulation is the ratio of two variance components $\lambda = \sigma_\epsilon^2 / \sigma_u^2$.

2.3 Bayesian penalized splines

In a Bayesian framework the penalty on spline coefficients is related to a specific prior distribution for $\boldsymbol{\theta}$. For example, a quadratic penalty $\boldsymbol{\theta}^t \boldsymbol{D\theta} / (2\sigma_\theta^2)$ is the special case of a Gaussian prior $\pi(\boldsymbol{\theta}) \propto \exp\{-\boldsymbol{\theta}^t \boldsymbol{D\theta} / (2\sigma_\theta^2)\}$, where the scaled penalty $\boldsymbol{D} / \sigma_\theta^2$ equals the precision matrix of the prior. Assuming normality for the responses Y_i, the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ for the spline coefficients under this prior is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \pi(\boldsymbol{Y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \propto \prod_{i=1}^n \exp\left[-\frac{1}{2\sigma_\epsilon^2} \{Y_i - \boldsymbol{P}(x_i, \underline{\tau}) \boldsymbol{\theta}\}^2 \right] \exp\left(-\frac{1}{2\sigma_\theta^2} \boldsymbol{\theta}^t \boldsymbol{D\theta} \right), \quad (5)$$

where $\pi(\boldsymbol{Y}|\boldsymbol{\theta})$ corresponds to the likelihood of the observation model (1). By taking logarithms and multiplying with $-2\sigma_\epsilon^2$, maximizing the posterior distribution in (5) is equivalent to minimizing $\sum_{i=1}^n \{Y_i - \boldsymbol{P}(x_i, \underline{\tau})\boldsymbol{\theta}\}^2 + \sigma_\epsilon^2 \boldsymbol{\theta}^t \boldsymbol{D}\boldsymbol{\theta}/\sigma_\theta^2$, so that the penalized spline estimator (3) and the posterior mode coincide for a fixed variance and smoothing parameter. Similar to the mixed model interpretation of penalized splines is that the smoothing parameter corresponds to the ratio of the error variance and the prior variance. The mixed model representation is a simple reparametrization of the Bayesian formulation of penalized splines that avoids the partial impropriety in the Gaussian prior if $\boldsymbol{D}$ is rank-deficient. Fahrmeir et al. (2004) employ this connection to derive empirical Bayes estimators based on mixed model methodology yielding posterior mode estimators.

In a fully Bayesian formulation, additional hyperpriors are assigned to the error variance σ_ϵ^2 and the prior variance σ_θ^2. The simplest and conjugate choices are inverse gamma distributions and a standard choice is $\sigma_\epsilon^2 \sim IG(0.001, 0.001)$ and $\sigma_\theta^2 \sim IG(0.001, 0.001)$. Inferences in the fully Bayesian approach are then typically based on Markov chain Monte Carlo (MCMC) simulation techniques, see Brezger and Lang (2006) for details.

3 Simultaneous Bayesian Credible Bands

In this section we focus on Bayesian credible bands derived from MCMC simulation output. In all approaches we assume that we are interested in computing simultaneous credible bands for a collection of function evaluations $\widehat{\boldsymbol{f}} = \boldsymbol{P}\hat{\boldsymbol{\theta}} = \{\hat{f}(x_1), \ldots, \hat{f}(x_n)\}^t$ based on simulation realizations $f^{(j)}(x_1), \ldots, f^{(j)}(x_n)$, $j = 1, \ldots, J$.

Note that the principle question in the construction of a Bayesian confidence band is conceptually different from frequentist confidence bands. The construction is based on the posterior distribution and one seeks a confidence region I_α such that $P_{\boldsymbol{f}|\boldsymbol{Y}}(\boldsymbol{f} \in I_\alpha) = 1-\alpha$, i.e. the coverage is defined in terms of the posterior distribution of $\boldsymbol{f} = \{f(x_1), \ldots, f(x_n)\}^t$

given the observed data $\boldsymbol{Y}$.

An obvious way to construct a simultaneous credible region for $\widehat{\boldsymbol{f}}$ is outlined in Crainiceanu et al. (2007). Suppose that $\widehat{\boldsymbol{f}}$ is the posterior mean estimator and that the posterior standard deviation $\sqrt{\widehat{\text{var}}\{\hat{f}(x_i)\}}$ for each point contained in $\widehat{\boldsymbol{f}}$ has been computed. By assuming approximate posterior normality and deriving the $(1-\alpha)$ sample quantile c_b of

$$\max_{i=1,\ldots,n} \left| \frac{f^{(j)}(x_i) - \hat{f}(x_i)}{\sqrt{\widehat{\text{var}}\{\hat{f}(x_i)\}}} \right|, \quad j = 1, \ldots, J, \tag{6}$$

a simultaneous credible region is given by the hyperrectangular

$$\left[\hat{f}(x_i) - c_b \sqrt{\widehat{\text{var}}\{\hat{f}(x_i)\}},\, \hat{f}(x_i) + c_b \sqrt{\widehat{\text{var}}\{\hat{f}(x_i)\}} \right], \quad i = 1, \ldots, n. \tag{7}$$

These confidence bands implicitly rely on the approximate normality. In particular, the standard deviation is used as a measure of uncertainty (assuming symmetry of the posterior distribution) and the posterior mean is considered as a center point. Hence, the full posterior distribution information contained in the sample is not utilized.

Alternatively, we propose a new simultaneous credible band that avoids the assumption of posterior normality but is still based on pointwise measures of uncertainty. To be more specific, we base our considerations on the pointwise credible intervals derived from the $\alpha/2$ and $1-\alpha/2$ quantiles of the samples $f^{(j)}(x_1), \ldots, f^{(j)}(x_n)$, $j = 1, \ldots, J$. In a second step, these pointwise credible intervals are scaled with a constant factor until $(1-\alpha)100\%$ of all sampled curves are contained in the credible band. The rationale is the following. The pointwise credible intervals provide us with a measure of where information on the estimated curve is sparse corresponding to wider intervals or dense corresponding to narrower intervals. In the approach by Crainiceanu et al. (2007) this information is obtained from posterior standard deviations. This, however, has the drawback that over- and underestimation of the penalized spline are treated in a symmetric fashion whereas the quantile-based approach allows for different uncertainty for over- and underestimation of the curve. This may be of particular relevance in local minima and maxima, where uncertainty may be attributed more strongly to one of the directions. While such differences may be generally small in Gaussian smoothing situations, they will typically become more relevant in non-Gaussian observation models. For example, the likelihood and therefore also the posterior in binary regression models or for Poisson data will be inherently asymmetric such that posterior normality will be questionable, at least for moderate sample sizes. In such situations, the proposed new band will be useful since it involves possibly asymmetric local measures of uncertainty. Note that determining Bayesian simultaneous confidence bands in non-Gaussian regression models is easy, since all computations only rely on the sampled curves and do not involve the simulation model. A further advantage over the credible band by Crainiceanu et al. (2007) is that our proposal does not depend on a specific point estimator, since our credible band makes full use of the posterior sample information, considering a $1-\alpha$ sample of the *curves* to determine the required scaling factor.

4 Simultaneous confidence bands with the volume of tube formula

4.1 The use of the volume of tube formula

The construction of simultaneous confidence bands using Weyl's (1939) volume of tube formula has been considered, among others, by Naiman (1986), Johansen and Johnstone (1990) and Sun and Loader (1994). While rigorous proofs are given by Sun (1993), we here sketch the basic ideas for completeness since these results will be used in the subsequent sections.

Let us consider the regression model (1) and some unbiased estimator $\tilde{f}(x) = \boldsymbol{l}(x)^t \boldsymbol{Y}$ with $\text{var}\{\tilde{f}(x)\} = \sigma_\epsilon^2 \|\boldsymbol{l}(x)\|^2$. Since $\tilde{f}(x)$ is unbiased, $Z(x) = \{\tilde{f}(x) - f(x)\}\sigma_\epsilon^{-1}\|\boldsymbol{l}(x)\|^{-1}$ is a zero mean Gaussian random field with $\text{var}\{Z(x)\} = 1$ and

$$\text{cov}\{Z(x_1), Z(x_2)\} = \left(\frac{\boldsymbol{l}(x_1)}{\|\boldsymbol{l}(x_1)\|}\right)^t \left(\frac{\boldsymbol{l}(x_2)}{\|\boldsymbol{l}(x_2)\|}\right) \equiv \sum_{i=1}^n v_i(x_1)v_i(x_2),$$

where $\sum_{i=1}^n v_i^2(x) = 1$. The set $V_n = \{\boldsymbol{v}(x) : x \in [a,b], \boldsymbol{v}(x) = (v_1(x), \ldots, v_n(x))\}$ is a one-dimensional manifold embedded in S^{n-1}, which is a unit sphere in $\mathbb{R}^n$. Let $\kappa_0 = \int_a^b \|\frac{d}{dx}\boldsymbol{v}(x)\|dx$ be the length of V_n and define the vector $\boldsymbol{\epsilon} = \boldsymbol{Y} - \boldsymbol{f}$. Then, Sun and Loader (1994) obtained that

$$\alpha = P\left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}(x)^t\boldsymbol{\epsilon}|}{\sigma_\epsilon\|\boldsymbol{l}(x)\|} \geq c\right) = \frac{\kappa_0}{\pi}\exp(-c^2/2) + 2\{1 - \Phi(c)\} + o\{\exp(-c^2/2)\}, \quad (8)$$

with $\Phi(\cdot)$ denoting the distribution function of a standard normal distribution. If σ_ϵ is unknown and is estimated with some $\hat{\sigma}_\epsilon$ such that $\varsigma\hat{\sigma}_\epsilon{}^2/\sigma_\epsilon^2 \sim \chi_\varsigma^2$, then

$$\alpha \approx \frac{\kappa_0}{\pi}\left(1 + \frac{c^2}{\varsigma}\right)^{-\varsigma/2} + P(|t_\varsigma| > c), \quad (9)$$

with t_ς a t-distributed random variable with ς degrees of freedom. A value for c is obtained from (9) and the simultaneous $100(1-\alpha)\%$ confidence band for $f(x)$ for x in the interval $[a,b]$ is constructed as

$$[\tilde{f}(x) - c\hat{\sigma}_\varepsilon\|\boldsymbol{l}(x)\|,\ \tilde{f}(x) + c\hat{\sigma}_\varepsilon\|\boldsymbol{l}(x)\|]. \quad (10)$$

4.2 Simultaneous confidence bands for penalized spline estimators

Consider now the penalized spline estimator with $\boldsymbol{l}(x) = \boldsymbol{P}(\boldsymbol{P}^t\boldsymbol{P} + \lambda\boldsymbol{D})^{-1}\boldsymbol{P}^t(x, \underline{\tau})$. In contrast to the setting of the previous section, $\boldsymbol{l}(x)$, as well as any other nonparametric estimator, is biased. A penalized spline estimator has two contributions to the bias. The approximation bias is due to the spline representation of the true function, while the shrinkage bias enters via the penalization. Theorem 1 of Claeskens et al. (2009) stated that depending on some assumptions on the number of knots K, the sample size n and the penalty λ, the theoretical properties of the penalized spline estimators are either similar to those of regression splines or to those of smoothing splines with a clear breakpoint between the two cases. In the later case, that is if the penalized splines asymptotics is close to that of smoothing splines, the shrinkage bias dominates the average mean squared error, while the approximation bias vanishes with the growing number of knots. Namely, the average squared approximation bias is of order $O(K^{-2q})$ with $K \sim \widetilde{C}\, n^{\nu/(2q+1)}$ for some constants $\widetilde{C}$ and $\nu > 1$, while the average squared shrinkage bias is of order $O\{n^{-2q/(2q+1)}\}$. Subsequently we assume that sufficiently many knots are taken, so that we can replace $f(x)$ with $\boldsymbol{P}(x, \underline{\tau})\boldsymbol{\theta}$ directly, with the approximation bias being negligible (see also assumption (A3) in the appendix).

Thus, for the construction of confidence bands one rather deals with

$$P_{\boldsymbol{Y}}\left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}(x)^t\boldsymbol{\epsilon} + m(x)|}{\sigma_\epsilon \|\boldsymbol{l}(x)\|} \geq c^*\right) = \alpha,$$

with $\boldsymbol{\epsilon} = \boldsymbol{Y} - \boldsymbol{P\theta}$, the shrinkage bias $m(x) = \boldsymbol{l}(x)^t\boldsymbol{P\theta} - \boldsymbol{P}(x, \underline{\tau})\boldsymbol{\theta}$ and a critical value c^* that accounts for the bias. The critical value c^* is typically difficult to find due to the unknown bias. Ignoring the shrinkage bias can lead to serious undercoverage, as is demonstrated in the simulation study presented in Section 6. Sun and Loader (1994) found that a plug-in correction with $m(x)$ replaced by an estimator fails badly, being in some cases even worse than no correction. They also suggested a bias correction procedure for a class of functions with Lipschitz continuous $m(x)/\|\boldsymbol{l}(x)\|$, based on the estimator of $\max_{x\in[a,b]} |m(x)|/\|\boldsymbol{l}(x)\|$. In their simulation study with the local polynomial regression estimates, the resulting coverage of the confidence bands appeared to be conservative and highly dependent on the choice of the smoothing parameter. Sun and Loader (1994) did not discuss a strategy for the best smoothing parameter choice in their setting. Clearly, choosing a smoothing parameter smaller than the optimal one in the mean squared error sense reduces the bias. However, no general guideline is available how small the smoothing parameter should be chosen. Note also that so far we assumed the smoothing parameter (λ or $\sigma_\epsilon^2/\sigma_u^2$) to be known. Replacing smoothing parameter by its estimator introduces an extra source of variability, which one has to account for.

In general, in this framework for penalized splines one faces the same problems as for any other nonparametric estimator – need for the bias correction and appropriate smoothing parameter choice. In the next section we consider simultaneous confidence bands which result from the mixed model representation of penalized splines and propose a simple bias correction for the standard nonparametric setting considered in this section.

4.3 Simultaneous confidence bands with the mixed model representation of penalized splines

4.3.1 Confidence bands based on the marginal mixed model

Let us now consider the mixed model representation of penalized splines, i.e. we approximate $f(x)$ by $\boldsymbol{P}(x,\underline{\tau})\boldsymbol{\theta} = \boldsymbol{X}(x)\boldsymbol{\beta} + \boldsymbol{Z}(x)\boldsymbol{u}$, with $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{\widetilde{K}})$ as in (4). Here $\boldsymbol{P}(x,\underline{\tau})\boldsymbol{\theta}$ is random due to randomness of $\boldsymbol{u}$. Note that Sun et al. (1999) worked with a similar mixed model, but they aimed to build a confidence bands around the marginal mean of $\boldsymbol{Y}$, that is around $\boldsymbol{X}(x)\boldsymbol{\beta}$ only. From the standard results on mixed models it is known that

$$Z_m(x) = \frac{\boldsymbol{P}(x,\underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\mathrm{var}\{\boldsymbol{P}(x,\underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}} = \frac{\boldsymbol{P}(x,\underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\sigma_\epsilon^2 \boldsymbol{P}(x,\underline{\tau})(\boldsymbol{P}^t\boldsymbol{P} + \sigma_\epsilon^2/\sigma_u^2 \boldsymbol{D})^{-1}\boldsymbol{P}(x,\underline{\tau})^t}} \sim N(0,1).$$

Since $\text{cov}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) = \sigma_\epsilon^2(\boldsymbol{P}^t\boldsymbol{P} + \sigma_\epsilon^2/\sigma_u^2\boldsymbol{D})^{-1}$, we find that $Z_m(x)$ is a nonsingular Gaussian zero mean random field with $\text{var}\{Z_m(x)\} = 1$ and

$$\text{cov}\{Z_m(x_1), Z_m(x_2)\} = \left(\frac{\boldsymbol{l}_m(x_1)}{\|\boldsymbol{l}_m(x_1)\|}\right)^t \left(\frac{\boldsymbol{l}_m(x_2)}{\|\boldsymbol{l}_m(x_2)\|}\right) \equiv \sum_{i=1}^{\widetilde{K}} v_{m,i}(x_1)v_{m,i}(x_2),$$

where $\boldsymbol{l}_m(x) = (\boldsymbol{P}^t\boldsymbol{P}+\sigma_\epsilon^2/\sigma_u^2\boldsymbol{D})^{-1/2}\boldsymbol{P}(x,\underline{\tau})^t$ is the $\widetilde{K}\times 1$ vector and $\boldsymbol{V}_{m,\widetilde{K}} = \{\boldsymbol{v}_m(x) : x \in [a,b], \boldsymbol{v}_m(x) = (v_{m,1}(x), \ldots, v_{m,\widetilde{K}}(x))\}$ is a one dimensional manifold embedded in $S^{\widetilde{K}-1}$. We replace κ_0 in (8) with the length of the mixed model manifold, $\kappa_{m,0} = \int_a^b \|\frac{d}{dx}\boldsymbol{v}_m(x)\|dx$, to obtain that

$$\begin{aligned}\alpha &= P_{\boldsymbol{Y},\boldsymbol{u}}\left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}_m(x)^t\boldsymbol{\epsilon}_m|}{\sigma_\epsilon\|\boldsymbol{l}_m(x)\|} \geq c_m\right) = P_{\boldsymbol{Y},\boldsymbol{u}}\left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}(x)^t\boldsymbol{\epsilon} + m(x,\boldsymbol{u})|}{\sigma_\epsilon\|\boldsymbol{l}_m(x)\|} \geq c_m\right) \quad (11)\\ &= \frac{\kappa_{m,0}}{\pi}\exp(-c_m^2/2) + 2\{1-\Phi(c_m)\} + o\{\exp(-c_m^2/2)\}, \quad (12)\end{aligned}$$

with $\boldsymbol{\epsilon}_m = (\boldsymbol{P}^t\boldsymbol{P} + \sigma_\epsilon^2/\sigma_u^2\boldsymbol{D})^{1/2}(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) \sim N(\boldsymbol{0}, \sigma_\epsilon^2\boldsymbol{I}_{\widetilde{K}})$. An unknown σ_ϵ can be replaced by any consistent estimator leading to an expression similar to (9).

Hence, our confidence band, obtained in the marginal mixed model framework is

$$[\tilde{f}_m(x) - c_m\hat{\sigma}_\epsilon\|\boldsymbol{l}_m(x)\|, \tilde{f}_m(x) + c_m\hat{\sigma}_\epsilon\|\boldsymbol{l}_m(x)\|]. \quad (13)$$

In practice, the smoothing parameter $\sigma_\epsilon^2/\sigma_u^2$ has to be replaced with its estimator. The following lemma shows that the variability due to smoothing parameter estimation can be ignored in the mixed model framework for n sufficiently large.

Lemma 1 *Under assumptions (A1)–(A3) listed in the appendix it holds*

$$\begin{aligned}\frac{\hat{\boldsymbol{l}}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})\boldsymbol{\theta}}{\|\hat{\boldsymbol{l}}_m(x)\|} &= \frac{\boldsymbol{l}_m(x)^t\boldsymbol{\epsilon}_m}{\|\boldsymbol{l}_m(x)\|} + O_p\left(n^{-\frac{1}{4q+2}}\right), \quad (14)\\ \frac{\hat{\boldsymbol{l}}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})\boldsymbol{\theta}}{\|\hat{\boldsymbol{l}}(x)\|} &= \frac{\boldsymbol{l}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})\boldsymbol{\theta}}{\|\boldsymbol{l}(x)\|} + O_p\left(n^{-\frac{1}{4q+2}}\right), \quad (15)\end{aligned}$$

with $\hat{\boldsymbol{l}}_m(x) = \boldsymbol{l}_m(x;\hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2)$ *and* $\hat{\boldsymbol{l}}(x) = \boldsymbol{l}(x;\hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2)$.

The proof is given in the appendix. Note that using a smaller q implies a smaller variability due to smoothing parameter estimation.

Using the same marginal mixed model framework for penalized splines, Ruppert et al. (2003) suggested a Monte Carlo procedure for estimation of c_m. Namely, a sufficiently large number ($N = 10,000$, say) of realizations of the random variable $(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}) \overset{\text{approx.}}{\sim} N(0, \hat{\sigma}_\epsilon^2(\boldsymbol{P}^t\boldsymbol{P} + \hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2\boldsymbol{D})^{-1})$ are generated and the corresponding values of

$$C = \max_{j=1,\ldots,M}\left[\frac{\boldsymbol{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})}{\sqrt{\widehat{\text{var}}\{\boldsymbol{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}}\right]$$

are calculated for a specified grid of x values $z_1, \ldots, z_M$. Their critical value $\hat{c}_m$ is the empirical $(1 - \alpha)$ quantile of the hence obtained values $C_1, \ldots, C_N$. A simultaneous confidence band is given by the hyperrectangular

$$\left[\boldsymbol{P}(z_j, \underline{\tau})\hat{\boldsymbol{\theta}}_m - \hat{c}_m\sqrt{\widehat{\text{var}}\{\boldsymbol{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}},\ \boldsymbol{P}(z_j, \underline{\tau})\hat{\boldsymbol{\theta}}_m + \hat{c}_m\sqrt{\widehat{\text{var}}\{\boldsymbol{P}(z_j, \underline{\tau})(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\}}\right], \quad (16)$$

for $j = 1, \ldots, M$. Note that this approach also does not take into account the variability due to variance parameters estimation. Hence, one can expect (13) and (16) to be approximately equal. Obviously, (7) is in fact the Bayesian version of (16), where the variability due to parameters estimation is taken into account. However, since in our simulation study in Section 6 we found no significant differences between the results obtained immediately from the tube formula and from (6), we believe that the tube formula offers an attractive alternative to the computationally intensive simulation based techniques.

4.3.2 Confidence bands based on the conditional mixed model

Let us now treat $\boldsymbol{u}$ in (4) as fixed and consider the probability

$$\alpha = P_{\boldsymbol{Y}|\boldsymbol{u}}\left(\max_{x\in[a,b]}\frac{\left|\boldsymbol{P}(x, \underline{\tau})(\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta})\right|}{\sigma_\epsilon\|\boldsymbol{l}(x)\|} \geq c^*\right) = P_{\boldsymbol{Y}|\boldsymbol{u}}\left(\max_{x\in[a,b]}\frac{|\boldsymbol{l}(x)^t\boldsymbol{\epsilon} + m(x, \boldsymbol{u})|}{\sigma_\epsilon\|\boldsymbol{l}(x)\|} \geq c^*\right), \quad (17)$$

where $\boldsymbol{l}(x) = \boldsymbol{l}(x, \sigma_\epsilon^2/\sigma_u^2)$. Up to a smoothing parameter this is exactly the probability discussed in Section 4.2. As already mentioned in Section 4.2, a plug-in correction of the form

$$\left[\tilde{f}(x) - \left(c + \frac{m(x,\boldsymbol{u})}{\sigma_\epsilon \|\boldsymbol{l}(x)\|}\right)\sigma_\epsilon \|\boldsymbol{l}(x)\|, \tilde{f}(x) + \left(c - \frac{m(x,\boldsymbol{u})}{\sigma_\epsilon \|\boldsymbol{l}(x)\|}\right)\sigma_\epsilon \|\boldsymbol{l}(x)\|\right], \tag{18}$$

with the bias $m(x, \boldsymbol{u})$ replaced by its estimate and c obtained from (8), performs poor. Instead, we suggest to use in place of c^* the critical value c_m obtained from (12). To justify this we compare (11) with (17) averaged over $\boldsymbol{u}$, that is with

$$\alpha = E_{\boldsymbol{u}} P_{\boldsymbol{Y}|\boldsymbol{u}} \left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}(x)^t \boldsymbol{\epsilon} + m(x,\boldsymbol{u})|}{\sigma_\epsilon \|\boldsymbol{l}(x)\|} \geq c^* \right) = P_{\boldsymbol{Y},\boldsymbol{u}} \left(\max_{x\in[a,b]} \frac{|\boldsymbol{l}(x)^t \boldsymbol{\epsilon} + m(x,\boldsymbol{u})|}{\sigma_\epsilon \|\boldsymbol{l}(x)\|} \geq c^* \right).$$

Hence,

$$P_{\boldsymbol{Y},\boldsymbol{u}} \left(\max_{x\in[a,b]} |Z_m(x)| \geq c_m \right) = P_{\boldsymbol{Y},\boldsymbol{u}} \left(\max_{x\in[a,b]} |Z_m(x)| \frac{\|\boldsymbol{l}_m(x)\|}{\|\boldsymbol{l}(x)\|} \geq c^* \right). \tag{19}$$

Formally, from (19) it follows that $c_m \leq c^* \leq r_m c_m$, with $r_m = \max_{x\in[a,b]} \|\boldsymbol{l}_m(x)\|/\|\boldsymbol{l}(x)\|$. To have an idea how big r_m can be, note that if K is small and $\lambda \to 0$, the penalized spline estimator converges to a projection (regression spline) estimator with $\boldsymbol{l}_m(x) \to \boldsymbol{l}(x)$, for all $x \in [a, b]$ so that $r_m \to 1$. As $K \to n$ a penalized spline estimator converges to a smoothing spline estimator, for which is known that $\sqrt{\text{ave}_x \|\boldsymbol{l}_m(x)\|^2/\text{ave}_x \|\boldsymbol{l}(x)\|^2} \approx \sqrt{2q/(2q-1)}$, see Wahba (1983). For example, for $q = 2$ one finds $\sqrt{2q/(2q-1)} \approx 1.15$.

Another way to look at c_m offers the following theorem.

Theorem 1 *For the critical values c and c_m, obtained from (8) and (12) respectively, it holds*

$$c_m^2 = c^2 + 2\frac{\kappa_{m,0} - \kappa_0}{\kappa_0} + o(1),$$

where $o(1)$ converges to zero as $c \to \infty$. Additionally, if

(A4) $\kappa_{m,0}{\kappa_0}^{-1} = \max_{x\in[a,b]} \|\boldsymbol{l}_m(x)\|^2 \|\boldsymbol{l}(x)\|^{-2} + o(1)$

is fulfilled and the mixed model (4) with $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{\tilde{K}})$ *holds, then*

$$c_m^2 = c^2 + \max_{x \in [a,b]} \frac{2 var_{\boldsymbol{u}}\{m(x, \boldsymbol{u})\}}{\sigma_\epsilon^2 \|\boldsymbol{l}(x)\|^2} + o(1). \tag{20}$$

The proof of the theorem can be found in the appendix and the assumption (A4) is easy to check in practice, using the R-package *ConfBands* that accompanies the paper. All together, we can conclude that the critical value c_m automatically accounts for the bias. Thus, one can build a confidence band

$$[\tilde{f}_m(x) - c_m \hat{\sigma}_\epsilon \|\boldsymbol{l}(x, \sigma_\epsilon^2/\sigma_u^2)\|, \tilde{f}_m(x) + c_m \hat{\sigma}_\epsilon \|\boldsymbol{l}(x, \sigma_\epsilon^2/\sigma_u^2)\|], \tag{21}$$

which will have approximately coverage probability $1 - \alpha$, given that enough knots are taken so that the approximation bias is negligible. Lemma 1 justifies replacement of the smoothing parameter by its estimate. In fact, the confidence band (21) is similar in spirit to the bias correction suggested by Sun and Loader (1994), but we avoid explicit estimation of $\max_{x \in [a,b]} |m(x, \boldsymbol{u})|/\|\boldsymbol{l}(x)\|$, using instead the critical value c_m obtained from the marginal mixed model framework.

5 Confidence bands in three frameworks

The confidence bands discussed in Sections 3 and 4 are obtained in different frameworks. They rely on different assumptions about the function f, the corresponding estimators use different smoothing parameter estimates and the interpretation of the confidence bands is also different. In the standard nonparametric model with f as a fixed sufficiently smooth function, the frequentist confidence bands are calculated with respect to the distribution of the data, given the function f. In other words, if one samples the data with the same mean function f many times, then one can expect that in $100(1 - \alpha)\%$ cases the true f will be inside the bands. In the Bayesian framework f is considered to be a sample path of a stochastic process and one is looking for the posterior probability that the true f is within the band, given the data. In the finite dimensional parametric setting both intervals – frequentist and Bayesian – are asymptotically equivalent. The well-known Bernstein-von Mises Theorem states that the posterior distribution of the finite dimensional parameter vector around its posterior mean is close to the distribution of the maximum likelihood estimate around the truth and herewith the Bayesian confidence sets have good frequentist coverage properties. Unfortunately, this is not true in the nonparametric regression context. In particular, Cox (1993) has shown that the Bayesian coverage probability for Bayesian smoothing splines with Gaussian priors tends to be larger than $(1-\alpha)100\%$, see also Freedman (1999). Hence, we expect to find our Bayesian credible bands to be conservative for $f \in C^q[a, b]$.

The mixed model based bands are something intermediate. On the one hand, one can consider them as an empirical version of the Bayesian confidence bands (with σ_ϵ, σ_u and $\boldsymbol{\beta}$ treated as fixed) having the same interpretation. On the other hand, one can view the mixed model based band as a confidence band averaged over $\boldsymbol{u}$. Thus, as shown in the previous section, the mixed model formulation of penalized splines can help to obtain confidence bands which have asymptotically either Bayesian or frequentist properties. Namely, the confidence band (13) is approximately equivalent to the Bayesian one and the band as defined in (21) has approximately frequentist properties. Our simulation results presented in Section 6 confirmed this.

The following theorem gives the asymptotic width of the intervals considered in our paper.

Theorem 2 *Under assumptions (A1)–(A3) the width of the confidence bands (10), (13) and (21) based on the volume of tube formula for a penalized spline estimator has the asymptotic order* $O_p(\sqrt{\log K^2} n^{-q/(2q+1)}) = O_p(\sqrt{\log n^{2\nu/(2q+1)}} n^{-q/(2q+1)})$, $\nu > 1$.

The proof is provided in the appendix. This theorem holds also if the smoothing parameter is replaced by its estimator $\hat{\sigma}_\epsilon^2/\hat{\sigma}_u^2$, as follows immediately from Lemma 1. It follows

that the interval is getting more narrow with growing n and getting wider with K.
Up to a constant $\sqrt{2\nu/(2q+1)}$ this asymptotic order coincides with the one given in Eubank and Speckman (1993) for a twice differentiable function ($q=2$), namely $O_p(\sqrt{\log n}n^{-2/5})$, which is slightly slower, than the optimal rate of $(\log n/n)^{q/(2q+1)}$ of Hall and Titterington (1988). Eubank and Speckman (1993) stressed that Hall and Titterington (1988) "chose a smoothing parameter designed to minimize the length of their intervals, rather than MSE" and conjectured that their rate of $\sqrt{\log n}n^{-2/5}$ is the best attainable with the smoothing parameter which minimizes the mean squared error. Using penalized splines one can get narrow intervals not only by taking a larger smoothing parameter, but also by choosing a smaller K. However, K should not be taken too small to avoid a growing approximation bias. More discussion on a practical choice of K is contained in Section 6.

6 Simulations

To assess the performance of the discussed approaches we ran a simulation study. We considered two functions. The first

$$f_1(x) = \frac{6}{10}\beta_{30,17}(x) + \frac{4}{10}\beta_{3,11}(x),$$

with $\beta_{l,m}(x) = \Gamma(l+m)\{\Gamma(l)\Gamma(m)\}^{-1}x^{l-1}(1-x)^{m-1}$ was used in Wahba (1983) and

$$f_2(x) = \sin^2\{2\pi(x-0.5)\}$$

has been considered in Eubank and Speckman (1993) and Xia (1998). These functions are shown in Figure 1. The x values are taken to be uniformly distributed over $[0,1]$. Three samples sizes were considered: a small one with $n=50$, a moderate one with $n=250$ and a large one with $n=500$. The errors are taken to be independent $N(0,\sigma_\epsilon^2)$ distributed with $\sigma_\epsilon = 0.3$. There are also simulation results available for $\sigma_\epsilon = 0.1$ and $\sigma_\epsilon = 0.6$,

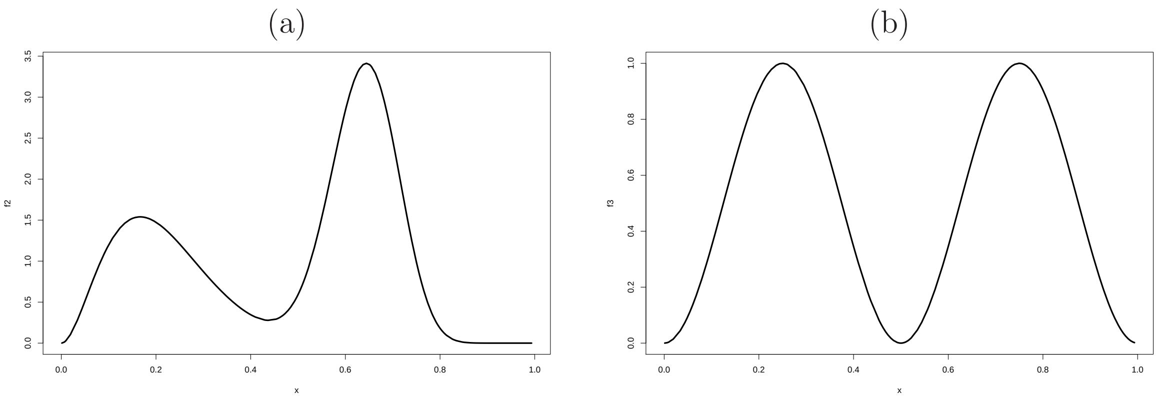

Figure 1: Functions used in the simulations: (a) $f_1(x)$, (b) $f_2(x)$.

but since there were no significant differences found we do not report them here. We estimated the curves with a different number of equidistant knots $K = 15$, 40, 100 and $K = 200$, depending on the sample size. Thereby we used a B-spline basis of degree 3 and as penalty the integrated squared second derivative of the spline function. The results for the 95% confidence bands that are reported in Table 1 are based on a Monte Carlo sample of size 1000.

The rows labeled $\boldsymbol{F}$ represent the coverage probabilities and corresponding areas for the confidence bands (10) built under the fixed effects nonparametric model without any bias correction, as described in Section 4.2. Since the volume of tube formula assumes the errors to be normally distributed but does not require $n \to \infty$, we do not discover any improvements in coverage with growing n, this holds for both functions. As expected from the results of Theorem 2, the width (and thus the area) of the bands is getting smaller as n increases. Overall we find that the confidence bands in the standard nonparametric framework which ignore the bias have on average a $5 - 10\%$ smaller coverage for all combinations of n and K.

The rows labeled $\boldsymbol{C}$ show the coverage probability of the mixed model based bands, conditional on $\boldsymbol{u}$, as discussed in Section 4.3.2. They should result in a coverage probability close to the nominal value, which we indeed observe. In general, it is recommended to

	$n = 50$		$n = 250$			$n = 500$			
	$K = 15$	40	$K = 15$	40	100	$K = 15$	40	100	200
f_1, $\boldsymbol{F}$	0.91	0.86	0.89	0.88	0.90	0.87	0.90	0.90	0.89
	(0.91)	(0.91)	(0.43)	(0.44)	(0.45)	(0.31)	(0.33)	(0.33)	(0.33)
$\boldsymbol{C}$	0.92	0.94	0.94	0.96	0.95	0.93	0.96	0.96	0.96
	(0.90)	(0.93)	(0.46)	(0.50)	(0.50)	(0.34)	(0.38)	(0.39)	(0.39)
$\boldsymbol{M}$	0.96	0.97	0.96	0.99	0.99	0.94	0.99	0.99	0.99
	(0.98)	(1.04)	(0.48)	(0.56)	(0.58)	(0.35)	(0.43)	(0.44)	(0.44)
$\boldsymbol{N}$	0.95	0.96	0.95	0.98	0.99	0.94	0.99	0.99	1.00
	(0.96)	(1.01)	(0.47)	(0.56)	(0.57)	(0.34)	(0.43)	(0.44)	(0.44)
$\boldsymbol{U}$	0.95	0.96	0.95	0.98	0.99	0.95	0.98	0.99	1.00
	(0.97)	(1.02)	(0.48)	(0.57)	(0.59)	(0.35)	(0.44)	(0.45)	(0.45)
f_2, $\boldsymbol{F}$	0.85	0.86	0.78	0.73	0.81	0.86	0.88	0.85	0.88
	(0.70)	(0.71)	(0.32)	(0.32)	(0.33)	(0.25)	(0.25)	(0.25)	(0.25)
$\boldsymbol{C}$	0.93	0.93	0.95	0.96	0.94	0.95	0.96	0.95	0.96
	(0.68)	(0.69)	(0.35)	(0.37)	(0.37)	(0.27)	(0.28)	(0.28)	(0.28)
$\boldsymbol{M}$	0.97	0.97	0.98	0.98	0.99	0.98	1.00	0.99	1.00
	(0.76)	(0.78)	(0.39)	(0.42)	(0.42)	(0.29)	(0.32)	(0.32)	(0.32)
$\boldsymbol{N}$	0.96	0.97	0.97	0.99	0.99	0.97	0.99	0.99	0.99
	(0.76)	(0.78)	(0.39)	(0.42)	(0.43)	(0.29)	(0.32)	(0.33)	(0.33)
$\boldsymbol{U}$	0.96	0.97	0.97	0.98	0.99	0.97	0.99	0.99	0.99
	(0.77)	(0.79)	(0.40)	(0.43)	(0.44)	(0.30)	(0.33)	(0.33)	(0.34)

Table 1: Coverage probabilities and (areas) for $f_1(x)$ and $f_2(x)$, with nominal level 0.95 using $\boldsymbol{F}$ a fixed effect model, $\boldsymbol{C}$ a mixed model conditional on $\boldsymbol{u}$, $\boldsymbol{M}$ a marginal mixed effect model and Bayesian method based on normal posteriors $\boldsymbol{N}$ and univariate credible bands $\boldsymbol{U}$. The range of the standard errors for the reported average areas is between 2.6% and 15.8% of the area for $f_1(x)$ and between 2.9% and 23.7% of the area for $f_2(x)$.

use moderate number of knots ($K = 25$ to 50, depending on the sample size) in this framework, since the width of the interval is growing with K. Overall, we find that the bias correction resulted from the mixed model representation of penalized splines is not only simple but is also very efficient.

The rows labeled $\boldsymbol{M}$ represent the coverage probabilities and corresponding areas for the confidence bands resulted from the marginal mixed model framework, as discussed in Section 4.3.1. These bands appear to become more and more conservative as K grows. This agrees with the finding of Cox (1993). Taking a small number of knots leads to a nearly parametric model where the smoothing parameter has little importance, which eliminates the differences between the mixed model representation of penalized splines and its standard nonparametric formulation. Thus, in the marginal mixed model framework taking a small K will imply less conservative bands from the frequentist point of view.

Finally, we consider the Bayesian confidence bands based on posterior normality (6) (denoted as $\boldsymbol{N}$). These bands are conceptually close to the marginal mixed model based bands, which is also reflected in a similar behaviour. This supports the asymptotic results of Lemma 1, which suggest that the variability due to smoothing parameter can be ignored. Our new proposed confidence bands denoted with $\boldsymbol{U}$ are typically somewhat wider than the $\boldsymbol{N}$ bands but in general yield a similar coverage, since the data were simulated from the normal distribution. Another method for the construction of Bayesian simultaneous credible bands can be found in Besag et al. (1995). This approach is based on order statistics of the samples. However, in our simulation study we found that the resulted credible bands suffer from serious undercoverage and we refrain on giving more details here.

Overall, we found that the frequentist confidence bands without any bias correction lead to undercoverage, while the Bayesian confidence bands typically become conservative with the growing K. The confidence bands based on the conditional mixed model result in the confidence bands with the coverage which is at most close to the nominal one.

7 Examples

To illustrate our method we present two examples. In the first example we consider the data on ratios of strontium isotopes found in fossil shells and their age, collected by T. Bralower of the University of North Carolina. First analyzed by Chaudhuri and Marron (1999), this data set was used in Section 6.2 on simultaneous confidence bands of Ruppert et al. (2003). 106 observations are shown in Figure 2. Ruppert et al. (2003)

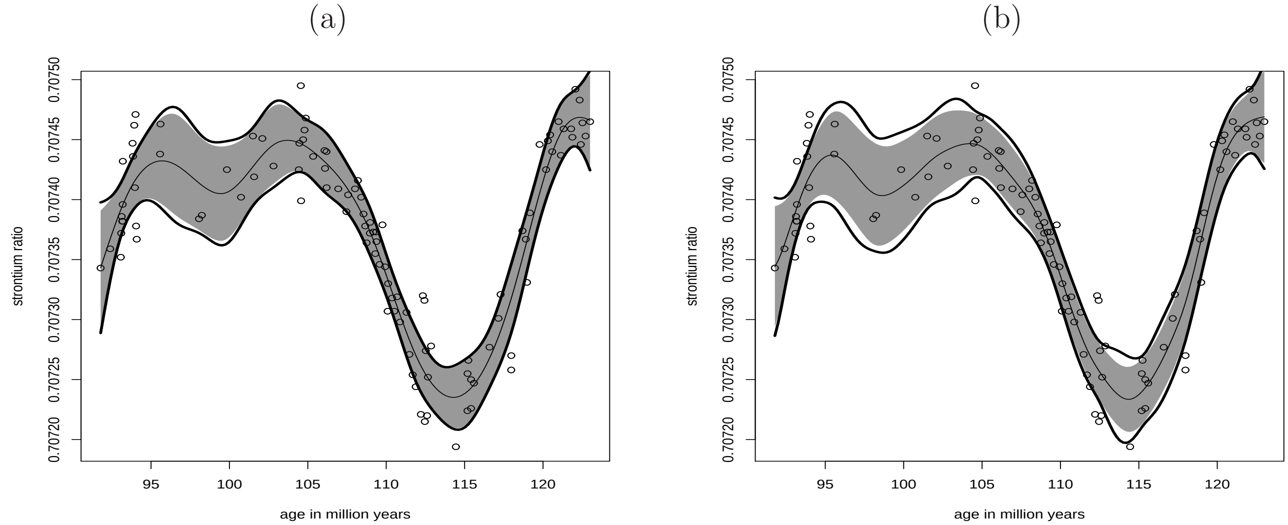

Figure 2: Fossil data with the confidence bands obtained from the conditional mixed model (shaded area) and the marginal mixed model (bold lines) (a) based on 10 knots, (b) based on 80 knots.

found the critical value using the simulation procedure described in Section 4.3.1, building the band (16). From five independent simulations of $N = 10,000$ each they got values $\hat{c}_m \simeq 3.172,\ 3.198,\ 3.172,\ 3.201,\ 3.199$.

From Theorem 2 follows that the critical values and thus the width of the band depends on the number of knots. We will illustrate this result using the fossil shells data. For the estimation we used penalized splines with the third degree B-splines basis and the second order penalty. We found critical values from the marginal mixed model using 10 and 80 knots. The volume of tube formula (12) delivered critical values $c_m(K = 10) = 3.229$ and $c_m(K = 80) = 3.380$. The simulation based critical values of Ruppert et al. (2003) for the band (16) with $M = 150$ and $N = 10,000$ in five independent runs resulted in $\hat{c}_m(K = 10) \simeq 3.107,\ 3.089,\ 3.080,\ 3.106,\ 3.092$ and to $\hat{c}_m(K = 80) \simeq 3.251,\ 3.267,\ 3.255,\ 3.253,\ 3.272$. Obviously, the growing number of knots results in somewhat bigger critical values. Note that Ruppert et al. (2003) did not provide the number of knots used, but one can conjecture it was around 30 – 40.

The grey area in Figure 2 represents the simultaneous confidence bands (21) obtained from the conditional mixed model and the bold lines show the bands (13) based on the marginal mixed model, using 10 knots (left) and 80 knots (right). The marginal mixed model based bands (16) and (13) with the critical values obtained from the simulations and with the volume of tube formula, respectively, are indistinguishable on the plot. As we already mentioned, the marginal mixed model based bands become more and more conservative as the number of knots grows, while the small number of knots leads to a nearly parametric model with a smaller difference between marginal and conditional mixed model based bands, which is clearly observed in Figure 2.

The second example is on undernutrition among children in Kenya. The data come from the 2003 Kenya Demographic and Health Survey (2003 KDHS) carried out by the Kenya Central Bureau of Statistics and are available free of charge from www.measuredhs.com for research purposes. We consider the so-called Z-score for stunting, depending on the

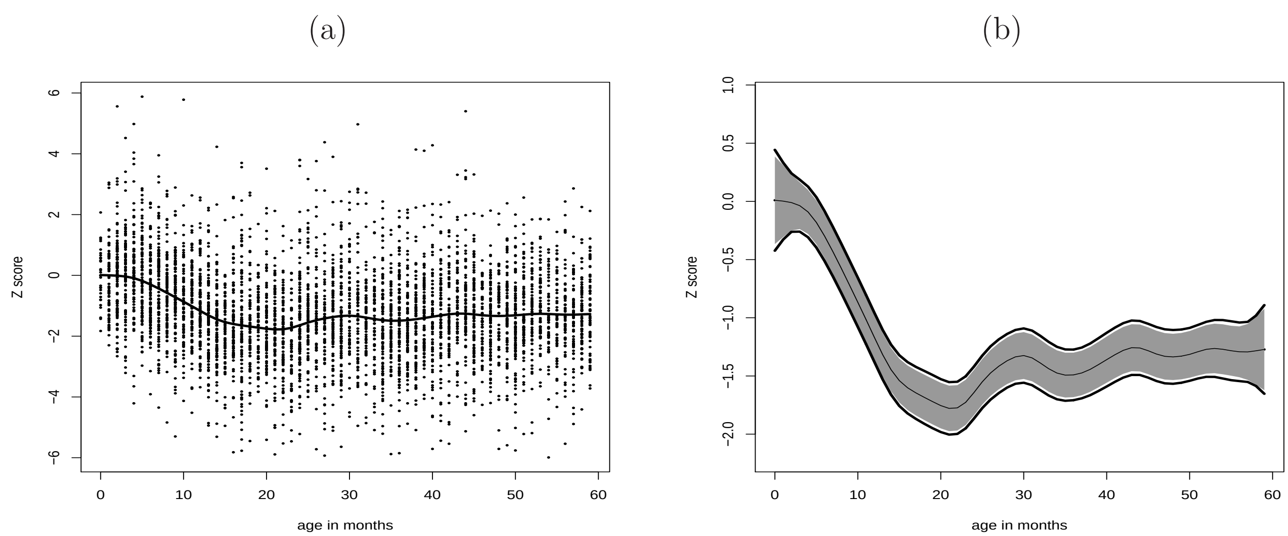

Figure 3: Z-scores for underweight of children in Kenya. (a) The data and a nonparametric fit based on 50 knots. (b) Simultaneous confidence bands based on the conditional mixed model (shaded area) and the marginal mixed model (bold lines).

age of a child. The Z-score for stunting is defined as the standardized height for age, i.e.

$Z = (H - m)/\sigma$, where H is the height of a child and m, σ are the median and the standard deviation of some reference population, correspondingly. The children with Z-scores below -2 are considered as stunted. All 4686 observations are shown in the left hand side plot of Figure 3. The data are cross-sectional (no same individuals) and available with the interval of one month. Note the low signal to noise ratio for this data. The grey area in the right hand side plot shows the confidence band (21) using the conditional mixed model representation of penalized splines based on 80 knots, while the bold lines are the corresponding marginal mixed model based bands (13), which are indistinguishable from the Bayesian confidence bands. Using these confidence bands one can perform a further analysis, e.g. a formal test for significance of bumps and dips between 1 and 4 years.

8 Discussion

In this paper we considered the construction of simultaneous confidence bands in three frameworks for penalized splines. We used the volume of tube formula in the standard nonparametric setting and for the mixed model representation of penalized splines. A full Bayesian analogue of the mixed model representation of penalized splines, as well as a new approach for Bayesian credible bands were considered. We found that the volume of tube formula for the mixed model formulation of penalized splines delivers results nearly identical to the full Bayesian framework, but with considerably less computational costs. Our main finding is that the mixed model formulation of penalized splines helps also to build the simultaneous bands with the frequentist coverage. Thereby no explicit bias estimation is necessary and the smoothing parameter is estimated from the corresponding (restricted) likelihood. Our approach appeared to be effective in the simulations, extremely fast and easy to implement. The R package *ConfBands* that accompanies the paper allows to obtain all the confidence bands discussed.

It is important to note that the volume of tube formula relies on the Gaussian distribution assumption for the errors. However, if the sample size is large and the central limit theorem applies, the volume of tube formula is still valid for the models with any non-Gaussian additive independent errors. Moreover, Loader and Sun (1997) showed that the volume of tube formula holds without modifications for spherically symmetric errors. In the linear regression context there were some modifications of the volume of tube formula developed in order to adjust for the cases with heteroscedastic (Faraway and Sun, 1995) and correlated errors (Sun et al., 1999), while Sun et al. (2000) considered generalized linear models. Extensions of our work to a generalized framework, as well as handling correlated and heteroscedastic data offers interesting directions for further research.

Acknowledgements

The authors are grateful to the editor, the referees of the paper and Dr. Tatiyana Apanasovich for the constructive remarks. We also thank Manuel Wiesenfarth for the substantial improvement of the R programme used for the Bayesian confidence bands. The first author acknowledges the support of the German Research Foundation (Deutsche Forschungsgemeinschaft) as part of the Institutional Strategy of the University of Göttingen.

Appendix. Technical details

A.1 Proofs

We adopt the framework of Claeskens et al. (2009) and use the same assumptions.

(A1) Let $\delta = \max_{0\le j\le K}(\delta_j)$, $\delta_j = \tau_j - \tau_{j-1}$. There exists a constant $M > 0$, such that $\delta / \min_{1\le j\le K}(\delta_j) \le M$ and $\max_{0\le j\le K} |\delta_{j+1} - \delta_j| = o(K^{-1})$.

(A2) For deterministic design points $x_i \in [a,b]$, $i = 1, \ldots, n$, assume that there exists a distribution function Q with corresponding positive continuous design density ρ such that, with Q_n the empirical distribution of $x_1, \ldots, x_n$, $\sup_{x\in[a,b]} |Q_n(x) - Q(x)| = o(K^{-1})$.

(A3) $K_q = (K + p + 1 - q)(\lambda\check{c})^{1/2q} n^{-1/2q} > 1$ for some constant $\check{c}$ that depends only on q and the design density ρ and $K \sim \widetilde{C}\, n^{\nu/(2q+1)}$ for some constants $\widetilde{C}$ and $\nu > 1$.

Proof of Lemma 1

Let us denote $\sigma_\epsilon^2/\sigma_u^2 = \lambda_m$. Since $\hat{\lambda}_m$ is a maximum likelihood estimator, a routine calculation shows that

$$\hat{\lambda}_m \overset{\text{approx.}}{\sim} N\left(\lambda_m, \frac{2\lambda_m^2}{\text{tr}(\boldsymbol{S}^2) - p + o(1)}\right),$$

where $o(1) \to 0$ as $n \to \infty$, $\text{tr}(\cdot)$ denotes the trace of the matrix and $\boldsymbol{S} = \boldsymbol{P}(\boldsymbol{P}^t\boldsymbol{P} + \lambda_m\boldsymbol{D})^{-1}\boldsymbol{P}^t$. We prove equation (15) only, the proof of (14) is completely analogous. Using

$$\frac{\partial\|\boldsymbol{l}(x)\|^{-1}}{\partial\lambda} = \frac{\boldsymbol{l}^t(x)(\boldsymbol{I}_n - \boldsymbol{S})\boldsymbol{l}(x)}{\lambda\|\boldsymbol{l}(x)\|^3}, \quad \frac{\partial\boldsymbol{l}(x)}{\partial\lambda} = \frac{(\boldsymbol{S} - \boldsymbol{I}_n)\boldsymbol{l}(x)}{\lambda},$$

and applying the delta method results in

$$\begin{aligned} \hat{\boldsymbol{l}}(x)^t\boldsymbol{f} &\overset{\text{approx.}}{\sim} N\left\{\boldsymbol{l}(x)^t\boldsymbol{f}, \text{var}(\hat{\lambda}_m)\left(\frac{\partial\boldsymbol{l}(x)^t\boldsymbol{f}}{\partial\lambda_m}\right)^2\right\}, \\ \|\hat{\boldsymbol{l}}(x)\|^{-1} &\overset{\text{approx.}}{\sim} N\left\{\|\boldsymbol{l}(x)\|^{-1}, \text{var}(\hat{\lambda}_m)\left(\frac{\partial\|\boldsymbol{l}(x)\|^{-1}}{\partial\lambda_m}\right)^2\right\}. \end{aligned}$$

With this one finds

$$\begin{aligned} \text{var}\left(\hat{\boldsymbol{l}}(x)^t\boldsymbol{f}\right) &= \frac{[\{(\boldsymbol{I}_n - \boldsymbol{S})\boldsymbol{l}(x)\}^t\boldsymbol{f}]^2}{\text{tr}(\boldsymbol{S}^2) - p + o(1)}, \\ \text{var}(\|\hat{\boldsymbol{l}}(x)\|^{-1}) &= \frac{\{\boldsymbol{l}^t(x)(\boldsymbol{I}_n - \boldsymbol{S})\boldsymbol{l}(x)\}^2}{2\|\boldsymbol{l}(x)\|^6\{\text{tr}(\boldsymbol{S}^2) - p + o(1)\}}. \end{aligned}$$

To obtain the asymptotic orders we use the results of Claeskens et al. (2009). In particular, from their Theorem 1 under assumptions (A1)–(A3) $n^{-1}\sum_{i=1}^{n}\text{var}\{\tilde{f}(x_i)\} = n^{-1}\text{tr}(\boldsymbol{S}^2) = O(n^{-2q/(2q+1)})$ for $K_q > 1$. Thus, $\text{var}\{\tilde{f}(x)\} = \sigma_\epsilon^2\|\boldsymbol{l}(x)\|^2 = O(n^{-2q/(2q+1)})$ for any $x \in [a, b]$ and $\text{tr}(\boldsymbol{S}^2) = O(n^{1/(2q+1)})$. With the arguments used in the proof of the asymptotic order of $\text{tr}(\boldsymbol{S}^2)$ in Claeskens et al. (2009), it is not difficult to see that $\text{tr}(\boldsymbol{S})$ and $\text{tr}(\boldsymbol{S}^3)$ have the same order $O(n^{1/(2q+1)})$, implying, in particular, that $\boldsymbol{l}(x)^t\boldsymbol{S}\boldsymbol{l}(x) = O(n^{-2q/(2q+1)})$. Noting that $\{(\boldsymbol{I}_n - \boldsymbol{S})\boldsymbol{l}(x)\}^t\boldsymbol{f} = \boldsymbol{l}^t(x)\{\boldsymbol{f} - E(\widetilde{\boldsymbol{f}})\}$, we conclude that its asymptotic order is the same as that of the bias of $\widetilde{\boldsymbol{f}}(x)$, that is $O(n^{-q/(2q+1)})$. Thus, we obtain $\hat{\boldsymbol{l}}(x)^t\boldsymbol{f} = \boldsymbol{l}(x)^t\boldsymbol{f} + O_p(n^{-1/2})$ and $\|\hat{\boldsymbol{l}}(x)\|^{-1} = \|\boldsymbol{l}(x)\|^{-1} + O_p(n^{(2q-1)/(4q+2)})$. Finally

$$\begin{aligned}
\frac{\hat{\boldsymbol{l}}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})^t\boldsymbol{\theta}}{\|\hat{\boldsymbol{l}}(x)\|} &= \frac{\boldsymbol{l}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})^t\boldsymbol{\theta} + \{\hat{\boldsymbol{l}}(x) - \boldsymbol{l}(x)\}^t\{\boldsymbol{f} + O_p(1)\}}{\|\boldsymbol{l}(x)\|}\,\frac{\|\boldsymbol{l}(x)\|}{\|\hat{\boldsymbol{l}}(x)\|} \\
&= \left\{\frac{\boldsymbol{l}(x)^t\boldsymbol{Y} - \boldsymbol{P}(x,\underline{\tau})^t\boldsymbol{\theta}}{\|\boldsymbol{l}(x)\|} + O_p\left(n^{-\frac{1}{4q+2}}\right)\right\}\left\{1 + O_p\left(n^{-\frac{1}{4q+2}}\right)\right\},
\end{aligned}$$

proving the lemma.

Proof of Theorem 1

From (8) and (12) we conclude

$$\frac{\kappa_{m,0}}{\pi}\exp(-c_m^2/2) - 2\Phi(c_m) + o\{\exp(-c_m^2/2)\} = \frac{\kappa_0}{\pi}\exp(-c^2/2) - 2\Phi(c) + o\{\exp(-c^2/2)\},$$

leading to $\exp(-c_m^2/2) = \exp(-c^2/2)\kappa_0\kappa_{m,0}{}^{-1}[1 + o\{\exp(-c^2/2)\}]$. Taking the logarithm from the both sides of the last equality and using the Taylor expansion of $\log(\kappa_{m,0})$ around $\log(\kappa_0)$, we find

$$c_m^2 = c^2 + 2\frac{\kappa_{m,0} - \kappa_0}{\kappa_0} + o(1),$$

where $o(1)$ converges to zero as $c \to \infty$. Note now that

$$\text{var}_{\boldsymbol{u}}\{m(x,\boldsymbol{u})\} = \text{var}_{\boldsymbol{u}}\{\boldsymbol{P}(x)(\boldsymbol{P}^t\boldsymbol{P} + \sigma_\epsilon^2/\sigma_u^2\boldsymbol{D})^{-1}\sigma_\epsilon^2/\sigma_u^2\boldsymbol{D}\boldsymbol{\theta}\} = \sigma_\epsilon^2(\|\boldsymbol{l}_m(x)\|^2 - \|\boldsymbol{l}(x)\|^2).$$

To obtain (20), it remains to apply (A4).

To prove Theorem 2 we need the following lemma.

Lemma 2 *Under assumption (A1) it holds*

$$\|\boldsymbol{l}_m^{'}(x)\|^2\|\boldsymbol{l}_m(x)\|^{-2} = c_1(x)K^2, \quad \{\boldsymbol{l}_m^t(x)\boldsymbol{l}_m^{'}(x)\}^2\|\boldsymbol{l}_m(x)\|^{-4} = \tilde{c}_1(x)K^2,$$

$$\|\boldsymbol{l}^{'}(x)\|^2\|\boldsymbol{l}(x)\|^{-2} = c_2(x)K^2, \quad \{\boldsymbol{l}^t(x)\boldsymbol{l}^{'}(x)\}^2\|\boldsymbol{l}(x)\|^{-4} = \tilde{c}_2(x)K^2,$$

for some positive constants $c_1(x)$*,* $c_2(x)$*,* $\tilde{c}_1(x)$*,* $\tilde{c}_2(x)$ *depending only on* p *and* x*.*

Proof of lemma 2

Without loss of generality we take p-degree (order $p+1$) B-splines as basis functions, so that for $x \in [\tau_i, \tau_{i+1})$, $i = 0, \ldots, K$ the basis vector takes the form $\boldsymbol{P}(x, \underline{\tau}) = \{\mathbf{0}_i, P_{i-p,p+1}(x), \ldots, P_{i,p+1}(x), \mathbf{0}_{K-i}\}$, with $P_{i,p+1}(x)$ denoting an ith B-spline of degree p evaluated at x and $\mathbf{0}_i$ as an i-dimensional vector of zeros. It is known (see e.g. Zhou et al., 1998) that $\boldsymbol{P}^{'}(x, \underline{\tau}) = p\boldsymbol{P}_p(x, \underline{\tau})\boldsymbol{\Delta}$, where $\boldsymbol{P}_p(x, \underline{\tau}) = \{\mathbf{0}_i, P_{i-p+1,p}(x), \ldots, P_{i,p}(x), \mathbf{0}_{K-i}\}$ and $\boldsymbol{\Delta}$ is a $(K+p) \times (K+p+1)$ matrix of weighted first order differences, that is a matrix with the rows $\{\mathbf{0}_j, -(\tau_{j+1+p} - \tau_{j+1})^{-1}, (\tau_{j+1+p} - \tau_{j+1})^{-1}, 0, \ldots\}$, $j = 0, \ldots, K+p-1$, each of length $K+p+1$. With this for $x \in [\tau_i, \tau_{i+1})$ we can rewrite $\|\boldsymbol{l}_m(x)\|^2 = \sum_{s,t=0}^{p} P_{i-s,p+1}(x)P_{i-t,p+1}(x)h_{st}$ and $\|\boldsymbol{l}(x)\|^2 = \sum_{s,t=0}^{p} P_{i-s,p+1}(x)P_{i-t,p+1}(x)\tilde{h}_{st}$, where $h_{st} = \{(\boldsymbol{P}^t\boldsymbol{P} + \lambda_m\boldsymbol{D})^{-1}\}_{i+1+s,i+1+t}$ and $\tilde{h}_{st} = \{(\boldsymbol{P}^t\boldsymbol{P} + \lambda_m\boldsymbol{D})^{-1}\boldsymbol{P}^t\boldsymbol{P}(\boldsymbol{P}^t\boldsymbol{P} + \lambda_m\boldsymbol{D})^{-1}\}_{i+1+s,i+1+t}$. Moreover,

$$\|\boldsymbol{l}_m^{'}(x)\|^2 = p^2 \sum_{s,t=0}^{p} h_{st} \left(\frac{P_{i-s,p}(x)}{\tau_{i+s+p} - \tau_{i+s}} - \frac{P_{i+1-s,p}(x)}{\tau_{i+1+s+p} - \tau_{i+1+s}} \right) \left(\frac{P_{i-t,p}(x)}{\tau_{i+t+p} - \tau_{i+t}} - \frac{P_{i+1+t,p}(x)}{\tau_{i+1+t+p} - \tau_{i+1+t}} \right),$$

$$\{\boldsymbol{l}_m^t(x)\boldsymbol{l}_m^{'}(x)\}^2 = p^2 \left\{ \sum_{s,t=0}^{p} h_{st} \left(\frac{P_{i-s,p}(x)}{\tau_{i+s+p} - \tau_{i+s}} - \frac{P_{i+1-s,p}(x)}{\tau_{i+1+s+p} - \tau_{i+1+s}} \right) P_{i-t,p}(x) \right\}^2,$$

where $P_{i-p,p}(x) = P_{i+1,p}(x) = 0$. The analogous expressions hold for $\|\boldsymbol{l}^{'}(x)\|$ and $\{\boldsymbol{l}^t(x)\boldsymbol{l}^{'}(x)\}^2$ with h_{st} replaced by $\tilde{h}_{st}$. According to (A1) there exist positive constants $c_{j,k} < \infty$ independent of n and K such that $(\tau_{j+k+p} - \tau_{j+k}) = c_{j,k}^{-1}\, p/K$. Thus, since $0 \leq P_{i,p+1}(x) \leq 1$ as well as $0 \leq P_{i,p}(x) \leq 1$ for any i and only h_{st} depends on K and n, we find that

$$c_1(x) = \frac{\sum_{s,t=0}^{p} h_{st} \left\{P_{i-s,p}(x)c_{i,s} - P_{i+1-s,p}(x)c_{i+1,s}\right\}\left\{P_{i-t,p}(x)c_{i,t} - P_{i+1-t,p}(x)c_{i+1,t}\right\}}{\sum_{s,t=0}^{p} h_{st}\; P_{i-s,p+1}(x)P_{i-t,p+1}(x)}$$

is independent of K and n. Similar expressions can be obtained for $\tilde{c}_1(x)$, $c_2(x)$ and $\tilde{c}_2(x)$.

Proof of Theorem 2

The width of the confidence band based on the volume of tube formula for penalized splines at a fixed x is determined by the critical value c or c_m and the standard deviation $\sigma_\epsilon \|\boldsymbol{l}(x)\|$ or $\sigma_\epsilon \|\boldsymbol{l}_m(x)\|$. From (8) and (12) follows that $c = \sqrt{\log[\kappa_0^2\{1 + O(1)\}]}$ and $c_m = \sqrt{\log[\kappa_{m,0}^2\{1 + O(1)\}]}$, where $O(1)$ is bounded for $c, c_m \to \infty$. As discussed in the proof of Lemma 1, the standard deviation $\sigma_\epsilon \|\boldsymbol{l}(x)\| = O(n^{-q/(2q+1)})$ and $\sigma_\epsilon \|\boldsymbol{l}_m(x)\| = O(n^{-q/(2q+1)})$. It remains to find the order of $\kappa_{m,0}$ and κ_0. By definition

$$\begin{aligned}
\kappa_{m,0} &= \int_a^b \left\| \frac{d}{dx} \frac{\boldsymbol{l}_m(x)}{\|\boldsymbol{l}_m(x)\|} \right\| dx = \int_a^b \frac{\sqrt{\|\boldsymbol{l}_m(x)\|^2 \|\boldsymbol{l}_m^{'}(x)\|^2 - \{\boldsymbol{l}_m(x)^t \boldsymbol{l}_m^{'}(x)\}^2}}{\|\boldsymbol{l}_m(x)\|^2} dx, \\
\kappa_0 &= \int_a^b \left\| \frac{d}{dx} \frac{\boldsymbol{l}(x)}{\|\boldsymbol{l}(x)\|} \right\| dx = \int_a^b \frac{\sqrt{\|\boldsymbol{l}(x)\|^2 \|\boldsymbol{l}^{'}(x)\|^2 - \{\boldsymbol{l}(x)^t \boldsymbol{l}^{'}(x)\}^2}}{\|\boldsymbol{l}(x)\|^2} dx.
\end{aligned}$$

Using lemma 2 we find $\kappa_{m,0} = O(K) = O(n^{\nu/(2q+1)})$, $\kappa_0 = O(K) = O(n^{\nu/(2q+1)})$ and the width of the confidence band based on the volume of tube formula for penalized splines has the asymptotic order $O_p(\sqrt{\log n^{2\nu/(2q+1)}}\, n^{-q/(2q+1)})$, $\nu > 1$.

References

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, 10:3–66.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1:1071–1095.

Brezger, A. and Lang, S. (2006). Generalized additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50:967–991.

Brumback, B. A., Ruppert, D., and Wand, M. P. (1999). Comment on Shively, Kohn and Wood. *J. Amer. Statist. Assoc.*, 94:794–797.

Chaudhuri, P. and Marron, J. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.*, 94(447):807–823.

Claeskens, G., Krivobokova, T., and Opsomer, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, 96(6):529–544.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, 31(6):1852–1884.

Cox, D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.*, 21(2):903–923.

Crainiceanu, C., Ruppert, D., Carroll, R., Adarsh, J., and Goodner, B. (2007). Spatially adaptive penalized splines with heteroscedastic errors. *J. Comput. Graph. Statist.*, 16:265–288.

Durban, M. and Currie, I. (2003). A note on p-spline smoothing with b-splines and penalties. *Comput. Statist.*, 11(2):89–121.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Stat. Science*, 11(2):89–121. With comments and a rejoinder by the authors.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.*, 88(424):1287–1301.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, 14:731–761.

Faraway, J. and Sun, J. (1995). Simultaneous confidence bands for linear regession with heteroscedastic errors. *J. Amer. Statist. Assoc.*, 90:1094–1098.

Freedman, D. (1999). On the Bernstein-von Mises Theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1140.

Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields*, 89(4):447–455.

Hall, P. and Opsomer, J. (2005). Theory for penalized spline regression. *Biometrika*, 92:105–118.

Hall, P. and Titterington, D. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.*, 27:228–254.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.*, 29(2):163–179.

Held, L. (2004). Simultaneous posterior probability statements from monte carlo output. *J. Comput. Graph. Statist.*, 13:20–35.

Hotelling, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Am. J. Math.*, 61:440–460.

Johansen, S. and Johnstone, I. M. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.*, 18(2):652–684.

Kauermann, G., Krivobokova, T., and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(2):487–503.

Knafl, G., Sacks, J., and Ylvisaker, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.*, 80(391):683–691.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, 95(2):415–436.

Loader, C. and Sun, J. (1997). Robustness of tube formula based confidence bands. *Journal of Computational and Graphical Statistics*, 6(2):242–250.

Naiman, D. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.*, 14:896–906.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Statist.*, 9(4):307–333.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Stat. Science*, 1:505–527. With discussion.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression.* Cambridge University Press, Cambridge, UK.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.*, 21(1):34–71.

Sun, J., Loader, C., and McCormick, W. (2000). Confidence bands in generalized linear models. *Ann. Statist.*, 28(2):429–460.

Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.*, 22(3):1328–1345.

Sun, J., Raz, J., and Faraway, J. J. (1999). Confidence bands for growth and response curves. *Statist. Sinica*, 9(3):679–698.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 45(1):133–150.

Weyl, H. (1939). On the volume of tubes. *Am. J. Math.*, 61:461–472.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4):797–811.

Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, 26(5):1760–1782.